

DOCUMENT RESUME

ED 402 326

TM 025 827

AUTHOR Beasley, T. Mark
TITLE Nonparametric Test of Discordance with Ordered Alternatives for "J" Groups.
PUB DATE 8 Apr 96
NOTE 25p.; Paper presented at the Annual Meeting of the American Educational Research Association (New York, NY, April 8-12, 1996).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)
EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *Comparative Analysis; *Group Membership; *Nonparametric Statistics; *Robustness (Statistics); Simulation
IDENTIFIERS Type I Errors

ABSTRACT

Robustness and power of parametric, semi-parametric, and nonparametric tests of between-group discordance were compared in this simulation study. The empirical Type I error rates and power of nine tests were compared. When data were sampled from the any differences especially favor young women in single-sex Catholic secondary schools, and whether pre-enrollment differences between students account for any sector nonnormality and a partial null hypothesis, the procedure of M. L. Puri and P. K. Sen (1985) and tests that use within-subject ranks were robust. The test developed by B. M. Katz and M. McSweeney (1983) showed erratic Type I error rates and was generally less powerful. A nonparametric extension of the L (M squared) of E. B. Page (1963) was robust and demonstrated consistent power functions. The paper details the development of M squared and provides computational examples. (Contains 9 tables and 29 references.) (Author/SLD)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

TM

Nonparametric Test of Discordance with Ordered Alternatives for J Groups

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

This document has been reproduced as
received from the person or organization
originating it.

Minor changes have been made to
improve reproduction quality.

Points of view or opinions stated in this
document do not necessarily represent
official OERI position or policy.

PERMISSION TO REPRODUCE AND
DISSEMINATE THIS MATERIAL
IN OTHER THAN PAPER COPY
HAS BEEN GRANTED BY

T. MARK BEASLEY

T. Mark Beasley
St. John's University, New York

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

American Educational Research Association, New York

April 8, 1996

ED 402 326

ABSTRACT: Robustness and power of parametric, semi-parametric, and nonparametric test of between group discordance were compared in this simulation study. When data were sampled from the exponential distribution, ranking of data was an effective in controlling Type I errors. Under combined conditions of nonnormality and a partial null hypothesis, the Puri & Sen (1985) procedure and tests that utilize within-subject ranks were robust. The Katz & McSweeney (1983) test showed erratic Type I error rates and was generally less powerful. A nonparametric extension of Page's (1963) L (M^2) was robust and demonstrated consistent power functions. The paper details the development of M^2 and provides computational examples.

Statistical procedures for making comparisons on the ranking of a set of k elements (e.g., treatments, preferences, alternatives) have been solved for the case in which the *ranking* is completed by a single group of n judges. In this situation, Friedman's (1937) test assesses the statistical significance of differences in mean rankings with the following null hypothesis:

$$H_0: \tau_1 = \tau_2 \dots = \tau_h \dots = \tau_k, \quad (1)$$

where τ_h represents the mean for the h^{th} treatment. The null hypothesis in this case assumes random rankings. This implies that for each judge, each of the $k!$ permutations of the $1, \dots, k$ ranks is equally likely to be assigned to the k alternatives.

In many experiments an *a priori* ordering of the k ranked elements can be expected from theoretical considerations and is thus of scientific interest. Whether these n judges agree with an *a priori* ordering of the k elements is assessed by an extension of the Friedman model, Page's (1963) L test of ordered alternatives.

BEST COPY AVAILABLE

TM 025827

Discordance Among J Groups

In practice, Page's L test is used to determine statistically whether the rankings one group of n judges agrees with an *a priori* ordering of k treatments. Each column of alternatives (T_h) is summed and weighted by the order of the treatment. Subsequently, these weighted components are summed to form L, which can be represented with the following formula:

$$L = \sum_{i=1}^n \sum_{h=1}^k hT_{ih}. \quad (2)$$

Thus, L when tested against the null hypothesis specified in (1) provides a nonparametric statistical index for the agreement of a group of n rankings to a set of k ordered alternatives (i.e., treatments, objects) with the following alternative hypothesis:

$$H_a: \tau_1 \leq \tau_2 \leq \dots \leq \tau_h \dots \leq \tau_k, \quad (3)$$

with at least one strict inequality (Page, 1963). However, this hypothesis is not tested exactly by the L test, in that rejection of the null hypothesis (1) does not indicate that H_a (3) is true. Therefore, Page's L, as well as the Friedman two-way ranked analysis of variance, tests the complete null hypothesis (1) against the most general alternative hypothesis which states that at least one pair among the k alternatives (τ_h) is not equal. Furthermore, the power of Page's L is an increasing function of the expected value of L. Therefore, Page's test could have substantial power even if none of the ranked treatments match the ordered alternatives specified by H_a . For example with $k = 4$ alternatives, if the true state of affairs is $\tau_1 < \tau_2 < \tau_3 < \tau_4$, Page's L (dependent on sample size, n), is likely to reject H_0 although H_a is not true (i.e., $\tau_2 < \tau_1 < \tau_4 < \tau_3$). Thus, Page's L provides an index to the degree of consensus to an *a priori* ranking.

Consistent with the Friedman model, the L test can be used to test linear and monotonic functions in the randomized block or repeated measures designs (Marascuilo & McSweeney, 1977). The L test of ordered hypotheses for multiple treatments is based on the linear contribution each ranking makes to the treatment sum of squares in the Friedman model or in a repeated measures analysis of variance (ANOVA) performed on the rankings. Thus, L was specifically designed for use with data measured at or reduced to an ordinal level (Page, 1963). It has also been shown to be equivalent to the average Spearman rank-order correlation between the *a priori* ordering and the group of n rankings (Lyerly, 1952; Page, 1963). That is, the average of the n correlations between individual rankings and the *a priori* ordering is mathematically related to L such that:

$$\bar{r} = \frac{12L}{nk(k^2 - 1)} - \frac{3(k+1)}{(k-1)}. \quad (4)$$

Given that (4) is a linear transformation of L (2), the averaged r could also be used as an index of concordance with ordered alternatives.

NONPARAMETRIC TEST OF DISCORDANCE AMONG GROUPS OF JUDGES

It should be noted that unlike the Mann-Whitney-Wilcoxon and Kruskal-Wallis tests, which are equivalent to a fixed-effect ANOVA with nested groups, designs employing a within-subject factor (i.e., the Friedman test and Page's L) do not have such straightforward nonparametric analogs. For two decades, the problem of concordance and discordance among J groups of judges has been addressed. The Schucany and Frawley (1973) statistic assesses the concordance in two groups in terms of the correlation between the mean of k ranks assigned by two independent groups of judges. It tests the null hypothesis of no between group concordance, regardless of within-group concordance. Rejections of the null hypothesis indicate between-group concordance. Beckett and Schucany (1979) developed a multiple comparison procedure for the Schucany-Frawley (SF) model; however, the issue of an *a priori* ordering of the elements is cannot be addressed.

Hollander and Sethuraman (1978) have questioned the SF model because it tests for positive rank correlation in mean ranks as an alternative hypothesis. Consistent with the Friedman model, the Hollander-Sethuraman (HS) model tests the identity of mean ranks across the k ranked elements as the null. Rejections of the null hypothesis indicate discordance between groups assuming within-group concordance. Serlin and Marascuilo (1983) developed planned and post-hoc comparison procedures for the HS model. Thus, to address the issue of an *a priori* ranking of the k elements (ordered alternatives in the Page test), linear contrasts could be imposed on the k ranked elements. However, these comparison methods only place linear restrictions on the between-groups contrasts and not on the coefficients applied to the k alternatives (within-subject factor). Therefore, there are an extremely large number of contrasts that could address the issue of ordered alternatives. Furthermore, the intercorrelation between the k ranked elements is not taken into account. Katz and McSweeney (1983) rectified these problems by developing omnibus, post-hoc, and linear contrast comparisons for the HS approach that uses the variance-covariance matrix for all rankings of the k elements. Thus, to test whether concordance with k ordered alternatives differs (i.e., discordance)

Discordance Among J Groups

across J groups each with n_j judges, a researcher should employ a valid interaction contrast (see Marascuilo & Levin, 1970) of the form:

$$\psi = \sum_{h=1}^k \sum_{j=1}^J a_{hj} \tau_{hj} \quad (5)$$

$$\text{where, } \sum_{h=1}^k a_{h*} = 0 \quad h = 1, \dots, k, \quad (6)$$

is a linear polynomial for k alternatives

$$\text{and } \sum_{j=1}^J a_{*j} = 0 \quad j = 1, \dots, J. \quad (7)$$

The variance of this contrast as elaborated by Marascuilo and McSweeney (1967) is

$$\text{Var}(\hat{\psi}) = \sum_{j=1}^J A_j V_j(\tau_k) A_j \quad (8)$$

where A_j is a $(k \times 1)$ linear contrast column vector for the j th group described in (6) and $V_j(\tau_k)$ is a $(k \times k)$ variance-covariance matrix for the j th group divided n_j (see Katz & McSweeney, 1983). Employing the Katz-McSweeney (KM) approach to test discordance among J groups, a test statistic, which approximates a chi-square distribution with $(J - 1)(k - 1)$ degrees-of-freedom (dfs), is formed by squaring (5) and dividing by (8).

EXTENSION OF PAGE'S L TO TEST DISCORDANCE AMONG J GROUPS

Another approach to the issue of whether J groups disagree in their consensus to an *a priori* ranking of k elements is to extend Page's L test for ordered alternatives. In such cases, three research questions can be asked. First, does the consensus of each group of judges agree with the *a priori* ranking of treatments? Second, does the consensus all judges agree with the *a priori* ranking of treatments? Third, do the groups differ in their degree of consensus with the ordered alternatives (*a priori* rankings). The first two questions are answered respectively by applying Page's L test to (a) each group of judges separately or (b) all judges regardless of group membership. In the two group situation, the third question can be answered by a simple extension of Page's L (Beasley & Leitner, 1993; Leitner & Dayton, 1976). Beasley (1993) proposed the M^2 test of discordance for multiple groups, factorial designs, and post-hoc comparisons. Since Page's L and the SF approach are both equivalent to the Spearman correlation under the condition of a single judge, both

are tests of mean rank correlation rather than mean rank identity (i.e., HS approach). Thus, to this extent the M^2 test is similar to the SF approach. The M^2 statistic, however, tests the complete null hypothesis (i.e., random rankings are assumed) that the J groups do not differ in their consensus with the ordered alternatives as measured by an "averaged L" statistic:

$$H_0: E(\bar{L}_1) = E(\bar{L}_2) = \dots = E(\bar{L}_J) \quad (9)$$

$$\text{where, } \bar{L}_j = \frac{L_j}{n_j}.$$

Rejection of the null leads to the conclusion that the groups are discordant in their consensus to the *a priori* ranking, which is similar to the HS approach. Therefore, the test of no differences in L statistics (9) can be employed to test the interaction of linear trends among the k ranks in a k by J split-plot design with a single repeated measure. That is, the L tests can be viewed as collapsing the repeated measures simple main effects into J statistics. Also, one overall L test ($\bar{L}_*$), collapsing group distinctions, would assess the ordered alternatives main effect; however, as in any crossed design the interaction (i.e., differences in main effects) is usually of most interest (Boik, 1993).

DEVELOPMENT OF THE M^2 TEST OF DISCORDANCE

Page (1963) showed that the L has an expected value of

$$E(L) = \frac{nk(k+1)^2}{4}. \quad (10)$$

and that under the complete null hypothesis, the L test has the following variance:

$$\text{Var}(L) = \frac{nk^2(k+1)^2(k-1)}{144}. \quad (11)$$

When testing discordance between J groups, the groups may be unequally sized. Thus, in scaling L for the j th group with a division by n_j , the expected value for the "averaged L" of the j th group is as follows:

$$E(\bar{L}_j) = \frac{k(k+1)^2}{4}. \quad (12)$$

Multiplying (11) by a constant $(1/n)$, the variance of $\bar{L}_j$ is

$$\text{Var}(\bar{L}_j) = \frac{nk^2(k+1)^2(k-1)}{144 n_j^2} = \frac{k^2(k+1)^2(k-1)}{144 n_j}. \quad (13)$$

Discordance Among J Groups

In order to form a statistic for testing discordance among J groups, let $\hat{\lambda}_j$ be an unbiased, independent, and asymptotically normally distributed estimate of the unknown population parameter that is indexed by $\bar{L}_j$. Also let $\hat{\lambda}_*$ be an estimate of the unknown common value of the $1, \dots, J$ parameters under the complete null hypothesis that they are identical which can be indexed by $\bar{L}_*$. Let $\text{Var}(\hat{\lambda}_j)$ be the large-sample estimate of the population variance. From this the χ^2 analog of Scheffe's (1959) theorem may be stated as:

$$U_*' = \sum_{j=1}^J \frac{(\hat{\lambda}_j - \hat{\lambda}_*)^2}{\text{Var}(\hat{\lambda}_j)}, \quad (14)$$

which approximates a chi-square distribution with $(J - 1)$ *dfs* (Marascuilo, 1966).

Thus to analyze the differences in L statistics, a formulation which uses deviations of "averaged sample L 's" ($\bar{L}_j$) from a "grand averaged L " ($\bar{L}_*$) can be developed with:

n_j = sample size for the j^{th} group,

J = number of groups,

N = total sample size, and

$$\bar{L}_* = \frac{1}{N} \sum_{j=1}^J L_j. \quad (15)$$

Assuming independent groups, the expected value of any group deviation squared is equal to the variance of $\bar{L}_j$ (13). In a one-way layout with J independent groups, deviations are squared, multiplied by their respective sample sizes, summed, and divided by a pooled within-group variance, $\text{Var}(\hat{\lambda}_j)$. Because each deviation from the mean can be considered a normal deviate, the summation of these deviates squared over J groups forms a statistic which approximates a chi-square distribution with $(J - 1)$ *dfs* which in keeping with (14) can be expressed as:

$$M_{(J-1)}^2 = \frac{144 \sum_{j=1}^J n_j (\bar{L}_j - \bar{L}_*)^2}{k^2(k+1)^2(k-1)}. \quad (16)$$

The M^2 test was developed on the assumption of the complete null hypothesis (i.e., random rankings within each group of judges) and has been shown to approximate a chi-square distribution with $(J - 1)$ *dfs* under those conditions (Beasley, 1993).

It is possible, however, for two groups of judges to agree with the *a priori* rankings but not differ from each other. Previous studies have shown that the

within-group consensus affects the variance of the Page's L , which in turn affects both the Type I error rate and power functions of the M^2 test (Beasley & Leitner, 1993). Therefore, the M^2 procedure may need modification so that a partial null hypothesis (i.e., between group differences without assuming random rankings) can be tested without affecting the Type I error rate and power of the tests. Because Page's L is linearly related to an average Spearman rank correlation (4), it can be transformed and used in a basic formula for computing the variance of a correlation:

$$\text{Var}(\bar{r}_j) = \frac{(1 - \bar{r}_j^2)^2}{(k - 1)}. \quad (17)$$

Given (4), the variance of the $\bar{L}_j$ is:

$$\text{Var}(\bar{L}_j) = \frac{k^2 (k^2 - 1)^2 (1 - \bar{r}_j^2)^2}{144(k - 1)n_j} = \frac{k^2 (k - 1) (k + 1)^2 (1 - \bar{r}_j^2)^2}{144 n_j} \quad (18)$$

As with correlations, the variance of Page's L decreases as it varies from its expected value. Therefore, when differences in the L statistics from J groups are tested (9) and these values deviate from the expected value, the within-group variance is lower than what is estimated in (16). Furthermore, when one group's L statistic is near the expected value and others are close to its maximum or minimum value, there may be extreme variance heterogeneity. Therefore, to modify the M^2 , a Welch (1951) type statistic was proposed. Because (18) involves several constants the following formulation is given:

$$M_W^2 = \frac{144 \sum_{j=1}^J c_j (\bar{L}_j - \tilde{L}) / (J - 1)}{k^2 (k + 1)^2 (k - 1) [1 + \frac{2(J-2)}{J^2 - 1} \sum_{j=1}^J (1 - \frac{c_j}{c^*})^2 / (n_j - 1)]} \quad (19)$$

$$\text{where, } c_j = \frac{n_j}{(1 - \bar{r}_j^2)^2}, \tilde{L} = \frac{\sum_{j=1}^J c_j \bar{L}_j}{c^*}, \text{ and } c^* = \sum_{j=1}^J c_j.$$

As compared to the M^2 , this Welch-type statistic is expected to approximate an F -distribution with degrees-of-freedom equal to $(J - 1)$ and

$$\left[\frac{3}{J^2 - 1} \sum_{j=1}^J (1 - \frac{c_j}{c^*})^2 / (n_j - 1) \right]^{-1}. \quad (20)$$

PARAMETRIC TESTS OF SPLIT-PLOT INTERACTION

In the examples given, judges are nested within a between-subjects factor and are crossed with the ordered alternatives (within-subject factor). The parametric technique used to analyze such research designs is the split-plot ANOVA which has the following model:

$$Y_{ijk} = \mu + \beta_j + \pi_{i(j)} + \gamma_k + \beta\gamma_{jk} + \gamma\pi_{ik(j)} + \epsilon_{ijk} . \quad (21)$$

Although more statistical power is generally provided, the within-subjects nature of the ordered alternatives in these designs also presents the disadvantage of an additional assumption concerning the form of the variance-covariance matrix. When variances and covariances are equal across levels of the within-subject factor, the matrix has compound symmetry which is a sufficient condition for the F statistic to be distributed as the F distribution under the null hypothesis (Huyhn & Feldt, 1970; Kirk, 1982). Approaches to handling violations of this assumption include rank transformations, the use of multivariate tests (Bock, 1975), and correcting degrees of freedom (Box, 1954).

The use of single df tests has also been proposed as an analytic approach to split-plot designs when parametric assumptions are violated (Boik, 1975, 1981; Rogan, Keselman, & Mendoza, 1979). Because Page's L and the proposed extensions convert the within-subjects factor with $(k - 1)$ dfs into a single statistic L , which is equivalent to the linear contribution to the Sum of Squares for ranks, the ordered alternative effect can be considered a single df test. Therefore, linear polynomial interaction contrasts (5) represent the difference in consensus with the k ordered alternative between J groups ($F_{(OD)}$). Therefore, statistical tests of these contrasts have $(J - 1)$ dfs .

RANK TESTS OF SPLIT-PLOT INTERACTION

Another popular alternative to ANOVA tests when parametric assumptions are violated is to rank the entire data matrix and either (a) apply the appropriate ANOVA model (linear interaction contrasts in this case) to the ranked data ($F_{(BG)}$) or (b) calculate the Puri and Sen (1985) statistic:

$$PS = (N^* - 1)\theta , \quad (22)$$

where N^* is the total number of observations and θ represents a measure of explained variation. In a regression context, θ is the rank analog of R^2 . In the split-plot ANOVA, PS can be calculated as:

$$PS = (N^* - 1) \text{RSS}(\text{linear interaction contrast}) / \text{RSS}(\text{total}), \quad (23)$$

where RSS refers to Rank Sum of Squares. Under conditions of the null hypothesis, PS approximates a chi-square distribution with $(J - 1)$ dfs .

Another approach would be to calculate within-subject rankings based on the Friedman model. Rasmussen (1989) demonstrated that as compared to the conventional F -test, various tests that use within-subject ranking of data in the split-plot layout controlled the Type I error rate under conditions of covariance heterogeneity. To perform within-subject ranking, the data for each individual subject is ranked separately. After within-subject ranking, the F -test for linear interactions from model (21) could be performed ($F_{(WS)}$). The Katz-McSweeney (KM) model with linear interaction contrasts (5 and 8) could be utilized. Furthermore, each individual's data could be formed into a L statistic (2). Analytic options include the one-factor (between-groups) ANOVA F -test ($F_{(L)}$), a Welch-test performed on the L statistics ($W_{(L)}$), the M^2 , or M^2_w .

COMPUTATIONAL EXAMPLE

Suppose an educational researcher randomly assigned high school science student teachers to $J = 2$ conditions. The treatment group ($n_2 = 6$) received specific training concerning the development of critical thinking skills for adolescents. The control group ($n_1 = 4$) received general lesson plans and reviewed lecture outlines with an experienced science teacher. The general research hypothesis was that the teaching styles of these student teachers would differ in the degree to which they were consistent with Bloom's (1956) taxonomy of educational objectives.

Videotapes of the student teachers lessons were reviewed and the percentage of time each teacher spent on activities related to $k = 5$ educational objectives (i.e., alternatives) was recorded. In Table 1, hypothetical data in percentages are cast into two-way tables with $k = 5$ columns of ordered alternatives (T_h) which represent ascending levels of Bloom's taxonomy: Knowledge (T_1); Comprehension (T_2); Application (T_3); Analysis (T_4); Synthesis (T_5). Because Bloom's taxonomy follows an increasing monotonic function, agreement with ordered alternatives as specified in hypothesis (3) were of interests. Whether the groups differed regarding this specific hypothesis is an issue of between group discordance.

The researcher expected that the Treatment Group would spend more time on activities as they related to higher levels of the cognitive domain. Hence, their teaching style would be in agreement (concordant) with Bloom's taxonomy. Whereas, the Control Group may distribute their teaching time much differently. Table 2 shows these data after within-subject ranking. Combined, Tables 1 and 2 give

Discordance Among J Groups

the information necessary to calculate the nine tests investigated in this study. All tests confirm that the teaching style of the Treatment group was consistent with Bloom's taxonomy more so than the style of the Control Group. That is, the Treatment group was significantly more concordant with the ordered alternatives (i.e., Bloom's taxonomy) than was the Control Group.

METHOD

Procedures

The empirical Type I error rates and Power of nine tests were compared using programs written in SAS/IML (SAS Institute, 1990). For the sake of simplicity, only the $J = 2$ group situation was simulated. Data were randomly generated separately for each case by generating a $(1 \times k)$ vector of random data. The consensus with the k ordered alternatives (ρ) was imposed by an algorithm developed by Beasley (1994) and collected over the sample size of each group (n_j). After the completion of the data generation for each group, the F -test for linear polynomial interaction contrasts (5) was performed on the original data ($F_{(OD)}$) using model (21). Then, the data were ranked over the entire data matrix irrespective of group distinctions and the F -test for linear polynomial interaction contrasts (5) was performed ($F_{(BG)}$) using model (21) and Puri-Sen (PS) test were performed. Subsequently, within-group ranking occurred and the F -test for linear polynomial interaction contrasts (5) was performed ($F_{(WS)}$) using model (21) and Katz-McSweeney (KM) test (eqs. 5 and 8) were performed. The within-subject rankings for each case were also transformed into L statistics and a between-subjects ANOVA F -test ($F_{(L)}$), Welch (1951) test ($W_{(L)}$), M^2 (16), and M^2_w (19) were performed. The data were collected over 5,000 replications and a statistical significance level of $\alpha = .05$.

Conditions

The number of alternatives were varied from $k = 3$ to 5. Data were generated from either the Normal or Exponential distributions using the RANNOR or RANEXP functions (SAS Institute, 1990), respectively. Total sample size was held constant at 20 and group sample size was varied under two conditions: equal samples of $n_j = 10$; or unequal samples of $n_1 = 5$ and $n_2 = 15$.

The consensus with the ordered alternatives (ρ_j) took on several conditions. For estimating empirical Type I error rates, the two groups had equal consensus of $\rho_j = 0.0, 0.4, 0.6$, and 0.8 . To investigate power of each test a difference in two group consensus was held constant at $(\rho_1 - \rho_2) = 0.4$. Seven power conditions were imposed.

In one condition ($\rho_1 = -0.2$; $\rho_2 = 0.2$) the difference in consensus is 0.4 and both groups had the same expected variance for L . In other conditions, the expected variances (see 17 and 18) are made disparate: ($\rho_1 = 0.0$; $\rho_2 = 0.4$); ($\rho_1 = 0.2$; $\rho_2 = 0.6$); ($\rho_1 = 0.4$; $\rho_2 = 0.8$). Because many of the tests being compared are sensitive to or control for variance heterogeneity, especially in presence of unequal sample sizes, these three conditions were reversed for disparate samples: ($\rho_1 = 0.4$; $\rho_2 = 0.0$); ($\rho_1 = 0.6$; $\rho_2 = 0.2$); ($\rho_1 = 0.8$; $\rho_2 = 0.4$). Although these conditions represent variance heterogeneity for the L statistics, no variance-covariance heterogeneity was imposed on the originally generated data in this study.

RESULTS

Type I Error Rate

Because empirical Type I error rates were estimated at the $\alpha = .05$ level of significance using 5,000 replications, the standard error of simulation was 0.003. Thus, values that exceeded 0.056 could be considered over 2 standard errors above α . However, almost 50 conditions were simulated; therefore, a considerable number of values could be expected to exceed 0.056 by chance. Thus, a general criterion of "values greater than 0.06," which is consistent with Bradley (1968), was employed as a guide for assessing robustness.

Tables 3, 4, and 5 show the empirical Type I error rates of all nine tests at the $\alpha = .05$ level of significance for $k = 3, 4$, and 5 alternatives, respectively. When groups are equally sized and are normally distributed, most tests hold the Type I error rate below α . However, the parametric F , $F(OD)$, seemed to inflate the Type I error rate as the within-group consensus increased to higher levels ($\rho_j = 0.8$). Ranking the data seemed to be an effective way to control Type I errors under condition of a partial null hypothesis (i.e., within-group consensus). However, except for $F(BG)$ and M_W^2 , the Type I error rate for these tests was reduced well below α , which may also affect their power. When sample sizes were unequal, the Welch procedure performed on L statistics, $W(L)$, and the newly proposed M_W^2 seriously inflated the Type I error rate even when the data were normally distributed. However, the robustness of M_W^2 seemed to improve as the number of alternatives increased from $k = 3$ to 5.

When the data were sampled from an exponential distribution, the Type I error rates of $F(OD)$ was seriously inflated under all partial null conditions. That is, $F(OD)$ was robust to distributional violations only when the complete null hypothesis ($\rho_j = 0$) was simulated. Ranking the entire data set ameliorates these problems with distributional assumptions as long as the Puri-Sen test was employed. Within-subject

Discordance Among J Groups

ranking also helped control Type I errors for data from the exponential distribution. Again, both $W(L)$ and M_W^2 demonstrated serious inflations with unequally sized samples. By contrast, the KM test had empirical Type I error rates drastically below α and thus appeared to be very conservative.

Therefore, the parametric F is adequately robust only when testing the complete null hypothesis for normally distributed data. For data sampled from the exponential distribution, however, ranking the data seems an appropriate and robust procedure. In the presence of a partial null hypothesis and nonnormal data, only the PS and tests that involve within-subject ranking are robust. Among the tests which employ within-subjects ranking, the ANOVA F performed on within-subject ranks ($F(WS)$) inflated the Type I error rate under partial null conditions. Also, the Type I error rates of the KM tests were erratic in that some inflations occurred when $k = 3$, but with $k = 5$, KM appeared conservative. Based on these results it appears that KM function is more robust with equal sample sizes or as the number of alternatives (k) increases. Thus, only PS , $F(L)$, and M^2 consistently held the of Type I error rate below α under all conditions. However, the KM and ANOVA F performed on within-subject ranks, $F(WS)$, could be considered competitors under combined conditions of nonnormality and a partial null hypothesis.

Power

Table 6 shows the comparative power for equal sample sizes of 10 for $k = 3, 4$, and 5 alternatives under the two distributional conditions. Tables 7, 8, and 9 show comparative power estimates for unequal sample sizes for $k = 3, 4$, and 5 alternatives, respectively. As can be seen in almost any table, for normally distributed data, parametric procedures are generally preferred, although M_W^2 is a strong competitor. When the normality assumption is violated the PS test or performing tests which utilize within-subjects rankings are most powerful. However, as k increased, PS became more powerful than the other tests.

Among the tests which utilize within-subject rankings, $F(WS)$ is the most powerful when there is considerable within-group consensus (i.e., $\rho_j = 0.8$). However, one must consider that $F(WS)$ inflated the Type I error rate under partial null conditions for data sampled from the exponential distribution. This abrogation became worse as the number of alternative increased to $k = 5$. Thus, $F(L)$, KM , and M^2 would be considered competitors among the Friedman (i.e., within-subject ranking) tests. However, both $F(L)$ and KM are subject to the problem of combining unequal sample sizes and unequal variances. $F(L)$ is more conservative when the smaller group had a

smaller expected variance (i.e., larger expected ρ_j). For example, when the smaller group ($n_1 = 5$) had an expected $\rho_1 = 0.4$ and the larger group ($n_2 = 15$) had an expected $\rho_2 = 0.8$ (which has a smaller expected variance than $\rho_1 = 0.4$, see eq. 18) the empirical power of $F(L)$ was .2146. By contrast, when the smaller group ($n_1 = 5$) had the smaller expected variance ($\rho_1 = 0.8$) and the larger group ($n_2 = 15$) had the larger expected variance ($\rho_2 = 0.40$), $F(L)$ had a considerably smaller power estimate of .1164. Interestingly, KM was even more influence by the combination of unequal sample sizes and variances, but it showed an opposite trend. That is, when the smaller group had the larger expected variance (smaller expected ρ_j), KM surprisingly demonstrated more power. This counterintuitive finding may be attributed to the estimation of separate covariance matrices in the KM test of specific hypotheses.

DISCUSSION

For split-plot designs which involve specific interaction hypotheses, conditions of nonnormality and partial null hypotheses are of definite concern. Under such conditions parametric procedures are not robust and the Puri-Sen rank transformation test controls the Type I error rate and provides a generally powerful approach. Within-subject ranking of the data also provides a robust although slightly less powerful alternative.

Thus, if the data are already ranked within each subject and concordance with ordered alternatives and discordance among groups is of interest, performing the split-plot ANOVA F with specific interaction hypotheses is the most powerful approach. However, this approach becomes less viable under conditions of nonnormality. Nonparametric extensions of the Friedman model and Page's L (i.e., M^2) provide tests which consistently hold the Type I error rate below the specified α and have more uniform power functions as compared to the Katz-McSweeney test for specific hypotheses and semi-parametric procedures such as $F(L)$. Specifically, the power functions of both $F(L)$ and KM were sensitive to combinations of unequal sample sizes and unequal expected variances. Unfortunately, M_W^2 and the Welch test performed on the L statistics were not robust in most situations where M^2 was less powerful. Thus, other formulations and approaches to testing discordance when variances or covariances differ across groups should be investigated. Furthermore, whether the within-subject ranking approach retains its power and robustness under these parametric violations for split-plot designs has yet to be determined. Based on the results of Rasmussen (1989) one would suspect that within-subject ranking may also be useful when the variance heterogeneity assumption is violated.

Discordance Among J Groups

However, the performance of these tests under covariance inequalities has yet to be investigated. Thus, future investigations should focus on how within-subject rank tests perform under conditions of sphericity and variance-covariance heterogeneity.

REFERENCES

- Beasley, T. M. (1993). *Multiple group and factorial rank test of consensus with ordered hypotheses: Extensions of Page's L test*. Unpublished doctoral dissertation. Southern Illinois University.
- Beasley, T. M. (1994). CORRMTX: Generating correlated data matrices in SAS/IML. *Applied Psychological Measurement: Computer Program Exchange*, 18, 95.
- Beasley, T. M., & Leitner, D. W. (1993, April). *Nonparametric test of ordered alternatives: Extensions of Page's L test for two groups of unequal size*. Paper presented at the meeting of American Educational Research Association. Atlanta, GA.
- Beckett, J., & Schucany, W. R. (1979). Concordance among categorical groups of judges. *Journal of Educational Statistics*, 4, 125-137.
- Bloom, B. S. (Ed.) (1956). *Taxonomy of educational objectives, handbook I: The cognitive domain*. New York: McKay.
- Bock, R. D. (1975). *Multivariate statistical methods in behavioral research*. New York: McGraw-Hill.
- Boik, R. J. (1993). The analysis of two-factor interactions in fixed effects linear models. *Journal of Educational Statistics*, 18, 1-40.
- Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems, I: Effect of inequality of variance in the one-way classification. *Annals of Mathematical Statistics*, 25, 290-302.
- Bradley, J. V. (1968). *Distribution-free statistical tests*. Englewood Cliffs, NJ: Prentice-Hall.
- Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, 32, 675-701.
- Hollander, M., & Sethuraman, J. (1978). Testing for agreement between two groups of judges. *Biometrika*, 65, 403-411.
- Huyhn, H., & Feldt, L. S. (1970). Conditions under which mean squares ratios in repeated measurements designs have exact F distributions. *Journal of the American Statistical Association*, 65, 1582-1585.
- Katz, B. M., & McSweeney, M. (1983). Some non-parametric test for analysing ranked data in multi-group repeated measures designs. *British Journal of Mathematical and Statistical Psychology*, 36, 145-156.
- Kirk, R. E. (1982). *Experimental design: Procedures for the behavioral sciences* (2nd ed.). Belmont CA: Brooks-Cole.
- Leitner, D. W., & Dayton, C. M. (1976, April). *A two-group test of ordered hypotheses for multiple treatments: An extension of Page's L test*. Paper presented at the meeting of the American Educational Research Association. San Francisco, CA.
- Lyerly, S. B. (1952). The average Spearman rank correlation coefficient. *Psychometrika*, 17, 412-428.
- Marascuilo, L. A. (1966). Large-sample multiple comparisons. *Psychological Bulletin*, 65, 280-290.

- Marascuilo, L. A., & Levin, J. R. (1970). Appropriate post hoc comparisons for interaction and nested hypotheses in analysis of variance designs. The elimination of Type IV errors. *American Educational Research Journal*, 7, 397-421.
- Marascuilo, L. A., & McSweeney, M. (1967). Nonparametric and post hoc comparisons for trend. *Psychological Bulletin*, 67, 401-412.
- Marascuilo, L. A., & McSweeney, M. (1977). *Nonparametric and distribution-free methods for the social sciences*. Monterey, CA: Brooks-Cole.
- Page, E. B. (1963). Ordered hypotheses for multiple treatments: A significance test for linear ranks. *Journal of American Statistical Association*, 58, 216-230.
- Puri, M. L., & Sen, P. K. (1985). *Nonparametric methods in general linear models*. New York: Wiley.
- Rasmussen, J. L. (1989). Parametric and non-parametric analysis of groups by trials design under variance-covariance inhomogeneity, *British Journal of Mathematical and Statistical Psychology*, 42, 91-102.
- Rogan, J. C., Keselman, H. L., & Mendoza, J. L. (1979). Analysis of repeated measurements. *British Journal of Mathematical and Statistical Psychology*, 32, 269-286.
- SAS Institute. (1990). *SAS/IML user's guide* (Release 6.04). Cary, NC: Author.
- Scheffe', H. (1959). *The analysis of variance*. New York: Wiley.
- Schucany, W. R., & Frawley, W. H. (1973). A rank test for two group concordance. *Psychometrika*, 38, 249-258.
- Serlin, R. C., & Marascuilo, L. A. (1983). Planned and post-hoc comparisons in tests of concordance and discordance for G groups of judges. *Journal of Educational Statistics*, 8, 187-206.
- Welch, B. (1951). On the comparison of several mean values: An alternative approach. *Biometrika*, 38, 330-336.

Discordance Among J Groups

Table 1. Computational Example, $J = 2$ groups, $k = 5$ alternatives.

Hypothetical Data in Percentages

Control Group	T ₁	T ₂	T ₃	T ₄	T ₅
	12	3	7	40	35
	10	30	41	16	2
	19	27	29	17	8
	18	4	22	20	17
Column Means	14.75	16.00	24.75	23.25	15.50
Treatment Group	T ₁	T ₂	T ₃	T ₄	T ₅
	16	1	8	30	34
	6	14	23	32	25
	5	8	15	29	43
	9	15	26	11	37
	3	10	20	28	38
	12	6	18	31	33
Column Means	8.50	9.00	18.33	26.83	35.00

Test of linear polynomial interaction contrasts, Model (21) $F(1, 32) = 9.31, p = .0046$.

Ranked Hypothetical Data

Control Group	T ₁	T ₂	T ₃	T ₄	T ₅
	17.50	3.50	9.00	48.00	45.00
	14.50	39.50	49.00	22.50	2.00
	28.00	35.00	37.50	24.50	11.00
	26.50	5.00	31.00	29.50	24.50
Column Means	21.625	20.750	31.625	31.125	20.625
Treatment Group	T ₁	T ₂	T ₃	T ₄	T ₅
	22.50	1.00	11.00	39.50	44.00
	7.50	19.00	32.00	42.00	33.00
	6.00	11.00	20.50	37.50	50.00
	13.00	20.50	34.00	16.00	46.00
	3.50	14.50	29.50	36.00	47.00
	17.50	7.50	26.50	41.00	43.00
Column Means	11.666	12.250	25.583	35.333	43.833

Test of linear polynomial interaction contrasts, Model (21) $F(1, 32) = 9.59, p = .0041$.

Puri-Sen statistic from (23),

$PS = 7.06, df = 1, p = .0077$.

Table 2. Computational Example, Hypothetical Data after Within-Subjects Ranking.

Control Group	T ₁	T ₂	T ₃	T ₄	T ₅	L _i
	3	1	2	5	4	51
	2	4	5	3	1	42
	3	4	5	2	1	39
	3	1	5	4	2	46
Column Means	2.75	2.50	4.25	3.50	2.00	44.50
Treatment Group	T ₁	T ₂	T ₃	T ₄	T ₅	L _i
	3	1	2	4	5	52
	1	2	3	5	4	54
	1	2	3	4	5	55
	1	3	4	2	5	52
	1	2	3	4	5	55
	2	1	3	4	5	54
Column Means	1.50	1.83	3.00	3.83	4.83	53.67

Test of linear polynomial interaction contrasts, Model (21), $F(1, 32) = 15.55, p = .0004$

ANOVA performed on the L statistics,

$$F(1, 8) = 17.86, p = .0029$$

Welch test performed on the L statistics,

$$F(1, 3.3) = 11.90, p = .0368$$

$$\text{From (16)} \quad M^2 = \frac{144[4(44.5-50)^2 + 6(53.67-50)^2]}{5^2(5+1)^2(5-1)} = 8.07,$$

$$M^2 = 8.07, df = 1, p = .0045$$

$$\text{From (4)} \quad \bar{r}_1 = -0.05 \quad \bar{r}_2 = 0.867,$$

$$\text{From (19)} \quad c_1 = 4.02, c_2 = 96.86, \tilde{L} = 53.3.$$

$$\text{From (20)} \quad dfd = 3.25$$

$$M_w^2(1, 3.25) = 12.97, p = .0322$$

Katz-McSweeney Approach

$$V_1(\tau_{k*}) = \begin{pmatrix} 0.0625 & -0.1250 & -0.0625 & 0.0417 & 0.0833 \\ -0.1250 & 0.7500 & 0.3750 & -0.5000 & -0.5000 \\ -0.0625 & 0.3750 & 0.5625 & -0.3750 & -0.5000 \\ 0.0417 & -0.5000 & -0.3750 & 0.4167 & 0.4167 \\ 0.0833 & -0.5000 & -0.5000 & 0.4167 & 0.5000 \end{pmatrix} \quad \Lambda'_1 = \{-2 \ -1 \ 0 \ 1 \ 2\},$$

$$V_2(\tau_{k*}) = \begin{pmatrix} 0.1167 & -0.0833 & -0.0667 & 0.0167 & 0.0167 \\ -0.0833 & 0.0944 & 0.0667 & -0.0722 & -0.0056 \\ -0.0667 & 0.0667 & 0.0667 & -0.0667 & 0.0000 \\ 0.0167 & -0.0722 & -0.0667 & 0.1611 & -0.0389 \\ 0.0167 & -0.0056 & 0.0000 & -0.0389 & 0.0278 \end{pmatrix} \quad \Lambda'_2 = \{2 \ 1 \ 0 \ -1 \ -2\},$$

$$\text{From (5)} \quad \psi^2 = 84.028$$

$$\text{From (8)} \quad \text{Var}(\hat{\psi}) = 7.0611,$$

$$KM = 11.90, df = 4, p = .0181$$

Discordance Among J GroupsTable 3. Empirical Type I Error Rates for Normal and Exponential Distributions, $k = 3$.

$k = 3$		$n_1 = 10$				$n_2 = 10$			
$\alpha = .05$		Normal				Exponential			
Test		$\rho_j = 0.0$	$\rho_j = 0.4$	$\rho_j = 0.6$	$\rho_j = 0.8$	$\rho_j = 0.0$	$\rho_j = 0.4$	$\rho_j = 0.6$	$\rho_j = 0.8$
$F(OD)$		0.0400	0.0496	0.0520	0.0644	0.0458	0.0800	0.0840	0.0888
$F(BG)$		0.0418	0.0486	0.0456	0.0506	0.0478	0.0690	0.0628	0.0602
PS		0.0452	0.0396	0.0250	0.0084	0.0500	0.0522	0.0392	0.0200
$F(WS)$		0.0430	0.0434	0.0388	0.0232	0.0452	0.0618	0.0526	0.0466
$F(L)$		0.0338	0.0352	0.0352	0.0316	0.0348	0.0416	0.0334	0.0344
$W(L)$		0.0444	0.0466	0.0432	0.0356	0.0440	0.0518	0.0452	0.0406
KM		0.0214	0.0250	0.0224	0.0176	0.0246	0.0284	0.0252	0.0234
M^2		0.0402	0.0370	0.0202	0.0056	0.0438	0.0470	0.0324	0.0200
M_w^2		0.0484	0.0562	0.0582	0.0492	0.0514	0.0714	0.0628	0.0664

$k = 3$		$n_1 = 5$				$n_2 = 15$			
$\alpha = .05$		Normal				Exponential			
Test		$\rho_j = 0.0$	$\rho_j = 0.4$	$\rho_j = 0.6$	$\rho_j = 0.8$	$\rho_j = 0.0$	$\rho_j = 0.4$	$\rho_j = 0.6$	$\rho_j = 0.8$
$F(OD)$		0.0482	0.0468	0.0518	0.0584	0.0488	0.0754	0.0856	0.0834
$F(BG)$		0.0492	0.0472	0.0448	0.0482	0.0516	0.0616	0.0634	0.0560
PS		0.0498	0.0378	0.0244	0.0082	0.0502	0.0492	0.0408	0.0180
$F(WS)$		0.0476	0.0398	0.0340	0.0256	0.0468	0.0566	0.0492	0.0364
$F(L)$		0.0384	0.0326	0.0318	0.0332	0.0410	0.0344	0.0322	0.0268
$W(L)$		0.0676	0.0820	0.0810	0.0648	0.0660	0.0796	0.0890	0.0772
KM		0.0582	0.0646	0.0574	0.0494	0.0582	0.0614	0.0640	0.0564
M^2		0.0492	0.0334	0.0208	0.0080	0.0476	0.0474	0.0340	0.0130
M_w^2		0.0702	0.0804	0.0960	0.1158	0.0686	0.0868	0.0974	0.1186

Table 4. Empirical Type I Error Rates for Normal and Exponential Distributions $k = 4$.

$k = 4$	$n_1 = 10$				$n_2 = 10$			
$\alpha = .05$	Normal				Exponential			
Test	$\rho_j = 0.0$	$\rho_j = 0.4$	$\rho_j = 0.6$	$\rho_j = 0.8$	$\rho_j = 0.0$	$\rho_j = 0.4$	$\rho_j = 0.6$	$\rho_j = 0.8$
$F(OD)$	0.0412	0.0430	0.0606	0.0722	0.0436	0.1018	0.0966	0.1126
$F(BG)$	0.0434	0.0412	0.0486	0.0532	0.0458	0.0746	0.0636	0.0670
PS	0.0490	0.0320	0.0272	0.0086	0.0496	0.0518	0.0380	0.0236
$F(WS)$	0.0450	0.0328	0.0292	0.0200	0.0494	0.0624	0.0588	0.0504
$F(L)$	0.0362	0.0326	0.0324	0.0312	0.0390	0.0344	0.0360	0.0336
$W(L)$	0.0506	0.0416	0.0440	0.0390	0.0512	0.0454	0.0454	0.0450
KM	0.0128	0.0100	0.0102	0.0088	0.0142	0.0130	0.0114	0.0114
M^2	0.0476	0.0276	0.0160	0.0034	0.0514	0.0450	0.0312	0.0152
M_W^2	0.0512	0.0472	0.0468	0.0536	0.0540	0.0584	0.0608	0.0748

$k = 4$	$n_1 = 5$				$n_2 = 15$			
$\alpha = .05$	Normal				Exponential			
Test	$\rho_j = 0.0$	$\rho_j = 0.4$	$\rho_j = 0.6$	$\rho_j = 0.8$	$\rho_j = 0.0$	$\rho_j = 0.4$	$\rho_j = 0.6$	$\rho_j = 0.8$
$F(OD)$	0.0506	0.0440	0.0580	0.0656	0.0452	0.1028	0.1028	0.1178
$F(BG)$	0.0504	0.0422	0.0428	0.0454	0.0426	0.0736	0.0698	0.0646
PS	0.0554	0.0358	0.0200	0.0064	0.0480	0.0572	0.0418	0.0224
$F(WS)$	0.0484	0.0376	0.0322	0.0190	0.0428	0.0652	0.0604	0.0438
$F(L)$	0.0396	0.0334	0.0348	0.0336	0.0330	0.0358	0.0364	0.0350
$W(L)$	0.0680	0.0694	0.0708	0.0704	0.0630	0.0706	0.0754	0.0780
KM	0.0346	0.0288	0.0326	0.0234	0.0298	0.0356	0.0340	0.0306
M^2	0.0540	0.0338	0.0188	0.0038	0.0472	0.0536	0.0382	0.0192
M_W^2	0.0560	0.0568	0.0702	0.0900	0.0490	0.0702	0.0788	0.0986

BEST COPY AVAILABLE

Discordance Among J GroupsTable 5 Empirical Type I Error Rates for Normal and Exponential Distributions $k = 5$.

$k = 5$		$n_1 = 10$				$n_2 = 10$			
$\alpha = .05$		Normal				Exponential			
Test	$\rho_j = 0.0$	$\rho_j = 0.4$	$\rho_j = 0.6$	$\rho_j = 0.8$	$\rho_j = 0.0$	$\rho_j = 0.4$	$\rho_j = 0.6$	$\rho_j = 0.8$	
$F(OD)$	0.0410	0.0436	0.0522	0.0780	0.0472	0.1130	0.1266	0.1346	
$F(BG)$	0.0430	0.0414	0.0414	0.0490	0.0440	0.0824	0.0820	0.0672	
PS	0.0454	0.0356	0.0210	0.0084	0.0498	0.0560	0.0448	0.0178	
$F(WS)$	0.0412	0.0384	0.0270	0.0138	0.0434	0.0750	0.0658	0.0426	
$F(L)$	0.0348	0.0372	0.0352	0.0304	0.0378	0.0364	0.0368	0.0316	
$W(L)$	0.0464	0.0494	0.0446	0.0382	0.0522	0.0488	0.0496	0.0394	
KM	0.0088	0.0068	0.0054	0.0050	0.0076	0.0084	0.0084	0.0040	
M^2	0.0472	0.0374	0.0158	0.0024	0.0488	0.0520	0.0414	0.0088	
M_W^2	0.0442	0.0496	0.0496	0.0474	0.0470	0.0638	0.0646	0.0650	

$k = 5$		$n_1 = 5$				$n_2 = 15$			
$\alpha = .05$		Normal				Exponential			
Test	$\rho_j = 0.0$	$\rho_j = 0.4$	$\rho_j = 0.6$	$\rho_j = 0.8$	$\rho_j = 0.0$	$\rho_j = 0.4$	$\rho_j = 0.6$	$\rho_j = 0.8$	
$F(OD)$	0.0416	0.0490	0.0520	0.0802	0.0488	0.1116	0.1256	0.1276	
$F(BG)$	0.0424	0.0430	0.0424	0.0546	0.0462	0.0782	0.0776	0.0706	
PS	0.0496	0.0352	0.0194	0.0100	0.0506	0.0538	0.0452	0.0214	
$F(WS)$	0.0442	0.0378	0.0248	0.0182	0.0442	0.0668	0.0652	0.0448	
$F(L)$	0.0348	0.0370	0.0316	0.0372	0.0400	0.0324	0.0362	0.0314	
$W(L)$	0.0656	0.0662	0.0698	0.0740	0.0634	0.0600	0.0742	0.0754	
KM	0.0190	0.0178	0.0166	0.0182	0.0222	0.0168	0.0236	0.0198	
M^2	0.0526	0.0334	0.0122	0.0030	0.0534	0.0492	0.0396	0.0146	
M_W^2	0.0444	0.0508	0.0584	0.0776	0.0460	0.0558	0.0740	0.0904	

Table 6. Comparative Power for Equal Sample Sizes of 10, $\alpha = .05$.

$k = 3$									
Normal					Exponential				
$n_1 = 10$ $n_2 = 10$	$\rho_1 = -0.2$ $\rho_2 = +0.2$	$\rho_1 = 0.0$ $\rho_2 = 0.4$	$\rho_1 = 0.2$ $\rho_2 = 0.6$	$\rho_1 = 0.4$ $\rho_2 = 0.8$	$\rho_1 = -0.2$ $\rho_2 = +0.2$	$\rho_1 = 0.0$ $\rho_2 = 0.4$	$\rho_1 = 0.2$ $\rho_2 = 0.6$	$\rho_1 = 0.4$ $\rho_2 = 0.8$	
$F(OD)$	0.1758	0.1816	0.1932	-----	0.1566	-----	-----	-----	
$F(BG)$	0.1642	0.1832	0.1866	0.2344	0.1342	-----	-----	-----	
PS	0.1672	0.1794	0.1618	0.1550	0.1206	0.0978	0.1294	0.1576	
$F(WS)$	0.1416	0.1476	0.1660	0.2178	0.1238	0.1158	0.1366	0.1900	
$F(L)$	0.1170	0.1248	0.1446	0.2058	0.0856	0.0876	0.0976	0.1410	
$W(L)$	0.1402	0.1508	0.1698	0.2300	0.1036	0.1068	0.1164	0.1612	
KM	0.0892	0.0942	0.1178	0.1624	0.0646	0.0666	0.0752	0.1134	
M^2	0.1368	0.1414	0.1420	0.1476	0.1110	0.1060	0.1128	0.1360	
M_{W}^2	0.1558	0.1692	0.1972	0.2654	0.1250	-----	-----	-----	
$k = 4$									
Normal					Exponential				
$n_1 = 10$ $n_2 = 10$	$\rho_1 = -0.2$ $\rho_2 = +0.2$	$\rho_1 = 0.0$ $\rho_2 = 0.4$	$\rho_1 = 0.2$ $\rho_2 = 0.6$	$\rho_1 = 0.4$ $\rho_2 = 0.8$	$\rho_1 = -0.2$ $\rho_2 = +0.2$	$\rho_1 = 0.0$ $\rho_2 = 0.4$	$\rho_1 = 0.2$ $\rho_2 = 0.6$	$\rho_1 = 0.4$ $\rho_2 = 0.8$	
$F(OD)$	0.2500	0.2684	-----	-----	0.2362	-----	-----	-----	
$F(BG)$	0.2444	0.2604	0.2838	0.3512	0.2014	-----	-----	-----	
PS	0.2588	0.2666	0.2508	0.2530	0.1766	0.1366	0.2018	0.2508	
$F(WS)$	0.2070	0.2220	0.2406	0.3318	0.1810	-----	0.2092	-----	
$F(L)$	0.1784	0.1924	0.2244	0.3258	0.1056	0.1236	0.1436	0.2122	
$W(L)$	0.2090	0.2252	0.2634	0.3646	0.1312	0.1504	0.1696	0.2500	
KM	0.0878	0.0976	0.1134	0.1680	0.0484	0.0588	0.0676	0.1076	
M^2	0.2142	0.2216	0.2128	0.2312	0.1568	0.1628	0.1732	0.2148	
M_{W}^2	0.2228	0.2434	0.2760	0.4070	0.1640	0.1754	-----	-----	
$k = 5$									
Normal					Exponential				
$n_1 = 10$ $n_2 = 10$	$\rho_1 = -0.2$ $\rho_2 = +0.2$	$\rho_1 = 0.0$ $\rho_2 = 0.4$	$\rho_1 = 0.2$ $\rho_2 = 0.6$	$\rho_1 = 0.4$ $\rho_2 = 0.8$	$\rho_1 = -0.2$ $\rho_2 = +0.2$	$\rho_1 = 0.0$ $\rho_2 = 0.4$	$\rho_1 = 0.2$ $\rho_2 = 0.6$	$\rho_1 = 0.4$ $\rho_2 = 0.8$	
$F(OD)$	0.3298	0.3500	0.3856	-----	0.3058	-----	-----	-----	
$F(BG)$	0.3146	0.3344	0.3742	0.4640	0.2558	-----	-----	-----	
PS	0.3354	0.3436	0.3392	0.3434	0.2224	0.1714	0.2640	0.3458	
$F(WS)$	0.2706	0.2898	0.3268	0.4408	0.2332	-----	-----	-----	
$F(L)$	0.2364	0.2496	0.3032	0.4458	0.1318	0.1476	0.1842	0.3016	
$W(L)$	0.2776	0.2932	0.3502	0.4942	-----	-----	-----	-----	
KM	0.0912	0.0938	0.1236	0.1890	0.0432	0.0454	0.0662	0.1128	
M^2	0.2960	0.3062	0.3122	0.3364	0.2058	0.2074	0.2404	0.3120	
M_{W}^2	0.2826	0.3098	0.3724	0.5396	0.1970	0.2054	-----	-----	

Note. Power estimates were omitted for conditions which inflated Type I error rates.

Discordance Among J Groups**Table 7.** Comparative Power for Normal and Exponential Distributions, unequal Sample Sizes, $k = 3$, $\alpha = .05$.

Normal							
$n_1 = 5$	$\rho_1 = -0.2$	$\rho_1 = 0.0$	$\rho_1 = 0.2$	$\rho_1 = 0.4$	$\rho_1 = 0.4$	$\rho_1 = 0.6$	$\rho_1 = 0.8$
$n_2 = 15$	$\rho_2 = +0.2$	$\rho_2 = 0.4$	$\rho_2 = 0.6$	$\rho_2 = 0.8$	$\rho_2 = 0.0$	$\rho_2 = 0.2$	$\rho_2 = 0.4$
$F(OD)$	0.1464	0.1666	0.1840	0.2472	0.1356	0.1426	0.1376
$F(BG)$	0.1372	0.1558	0.1876	0.2456	0.1346	0.1396	0.1386
PS	0.1384	0.1470	0.1470	0.1344	0.1372	0.1336	0.1046
$F(WS)$	0.1172	0.1284	0.1588	0.2210	0.1190	0.1112	0.1060
$F(L)$	0.1002	0.1108	0.1460	0.2308	0.0990	0.0910	0.0842
$W(L)$	-----	-----	-----	-----	-----	-----	-----
KM	0.1130	-----	0.0824	-----	-----	0.2122	-----
M^2	0.1162	0.1226	0.1304	0.1286	0.1206	0.1060	0.0832
M_W^2	-----	-----	-----	-----	-----	-----	-----

Exponential							
$n_1 = 5$	$\rho_1 = -0.2$	$\rho_1 = 0.0$	$\rho_1 = 0.2$	$\rho_1 = 0.4$	$\rho_1 = 0.4$	$\rho_1 = 0.6$	$\rho_1 = 0.8$
$n_2 = 15$	$\rho_2 = +0.2$	$\rho_2 = 0.4$	$\rho_2 = 0.6$	$\rho_2 = 0.8$	$\rho_2 = 0.0$	$\rho_2 = 0.2$	$\rho_2 = 0.4$
$F(OD)$	0.1436	-----	-----	-----	-----	-----	-----
$F(BG)$	0.1232	-----	-----	-----	-----	-----	-----
PS	0.1140	0.1104	0.1168	0.1340	0.0516	0.1114	0.1174
$F(WS)$	0.1084	0.1126	0.1384	0.1918	0.0864	0.1120	0.1210
$F(L)$	0.0726	0.0856	0.1058	0.1602	0.0668	0.0706	0.0736
$W(L)$	-----	-----	-----	-----	-----	-----	-----
KM	0.0874	-----	-----	-----	-----	-----	-----
M^2	0.0978	0.1016	0.1094	0.1284	0.0838	0.0976	0.0926
M_W^2	-----	-----	-----	-----	-----	-----	-----

Note. Power estimates were omitted for conditions which inflated Type I error rates.

Table 8. Comparative Power for Normal and Exponential Distributions, unequal Sample Sizes, $k = 4$, $\alpha = .05$.

Normal							
$n_1 = 5$	$\rho_1 = -0.2$	$\rho_1 = 0.0$	$\rho_1 = 0.2$	$\rho_1 = 0.4$	$\rho_1 = 0.4$	$\rho_1 = 0.6$	$\rho_1 = 0.8$
$n_2 = 15$	$\rho_2 = +0.2$	$\rho_2 = 0.4$	$\rho_2 = 0.6$	$\rho_2 = 0.8$	$\rho_2 = 0.0$	$\rho_2 = 0.2$	$\rho_2 = 0.4$
$F(OD)$	0.1982	0.2128	0.2698	-----	0.2020	0.2028	-----
$F(BG)$	0.1910	0.2084	0.2684	0.3302	0.2092	0.1886	0.1630
PS	0.2012	0.2016	0.2094	0.2002	0.1372	0.1336	0.1046
$F(WS)$	0.1642	0.1806	0.2308	0.2962	0.1656	0.1574	0.1776
$F(L)$	0.1406	0.1606	0.2176	0.3248	0.1382	0.1354	0.1526
$W(L)$	-----	-----	-----	-----	-----	-----	-----
KM	0.1046	0.0826	0.0888	0.0914	0.1438	0.1764	0.2694
M^2	0.1770	0.1770	0.1934	0.1936	0.1780	0.1604	0.1470
M_W^2	0.1580	0.1290	-----	-----	0.2098	-----	-----

Exponential							
$n_1 = 5$	$\rho_1 = -0.2$	$\rho_1 = 0.0$	$\rho_1 = 0.2$	$\rho_1 = 0.4$	$\rho_1 = 0.4$	$\rho_1 = 0.6$	$\rho_1 = 0.8$
$n_2 = 15$	$\rho_2 = +0.2$	$\rho_2 = 0.4$	$\rho_2 = 0.6$	$\rho_2 = 0.8$	$\rho_2 = 0.0$	$\rho_2 = 0.2$	$\rho_2 = 0.4$
$F(OD)$	0.2056	-----	-----	-----	-----	-----	-----
$F(BG)$	0.1654	-----	-----	-----	-----	-----	-----
PS	0.1484	0.1346	0.1696	0.1956	0.0802	0.1596	0.1948
$F(WS)$	0.1522	-----	-----	-----	-----	-----	-----
$F(L)$	0.0880	0.0950	0.1346	0.2146	0.1010	0.0922	0.1164
$W(L)$	-----	-----	-----	-----	-----	-----	-----
KM	0.0662	0.0618	0.0602	0.0680	0.0930	0.1168	0.1946
M^2	0.1378	0.1286	0.1552	0.1820	0.1330	0.1386	0.1604
M_W^2	0.1276	0.0958	-----	-----	0.1584	-----	-----

Note. Power estimates were omitted for conditions which inflated Type I error rates.

BEST COPY AVAILABLE

Discordance Among J Groups**Table 9.** Comparative Power for Normal and Exponential Distributions, unequal Sample Sizes, $k = 5$, $\alpha = .05$.

Normal							
$n_1 = 5$	$\rho_1 = -0.2$	$\rho_1 = 0.0$	$\rho_1 = 0.2$	$\rho_1 = 0.4$	$\rho_1 = 0.4$	$\rho_1 = 0.6$	$\rho_1 = 0.8$
$n_2 = 15$	$\rho_2 = +0.2$	$\rho_2 = 0.4$	$\rho_2 = 0.6$	$\rho_2 = 0.8$	$\rho_2 = 0.0$	$\rho_2 = 0.2$	$\rho_2 = 0.4$
$F(OD)$	0.2584	0.2816	0.3424	-----	0.2660	0.2672	-----
$F(BG)$	0.2480	0.2730	0.3416	0.4370	0.2538	0.2596	0.3046
PS	0.2626	0.2630	0.2792	0.2780	0.2718	0.2594	0.2460
$F(WS)$	0.0594	0.2622	0.6886	0.9834	0.2158	0.2296	0.2752
$F(L)$	0.1882	0.2180	0.2900	0.4426	0.1904	0.1988	0.2500
$W(L)$	-----	-----	-----	-----	-----	-----	-----
KM	0.1042	0.0948	0.0898	0.1094	0.1368	0.1842	0.3018
M^2	0.2334	0.2432	0.2466	0.2722	0.2312	0.2334	0.2236
M_W^2	0.1894	0.1594	0.1412	-----	0.2452	0.3416	-----

Exponential							
$n_1 = 5$	$\rho_1 = -0.2$	$\rho_1 = 0.0$	$\rho_1 = 0.2$	$\rho_1 = 0.4$	$\rho_1 = 0.4$	$\rho_1 = 0.6$	$\rho_1 = 0.8$
$n_2 = 15$	$\rho_2 = +0.2$	$\rho_2 = 0.4$	$\rho_2 = 0.6$	$\rho_2 = 0.8$	$\rho_2 = 0.0$	$\rho_2 = 0.2$	$\rho_2 = 0.4$
$F(OD)$	0.2670	-----	-----	-----	-----	-----	-----
$F(BG)$	0.2246	-----	-----	-----	-----	-----	-----
PS	0.1906	0.1738	0.2136	0.2692	0.1048	0.2138	0.2696
$F(WS)$	0.0584	-----	-----	-----	-----	-----	-----
$F(L)$	0.1118	0.1326	0.1710	0.2836	0.1232	0.1256	0.1716
$W(L)$	-----	-----	-----	-----	-----	-----	-----
KM	0.0568	0.0648	0.0540	0.0788	0.0770	0.1078	0.1896
M^2	0.1734	0.1726	0.2000	0.2414	0.1626	0.1896	0.2198
M_W^2	0.1422	-----	-----	-----	-----	-----	-----

Note. Power estimates were omitted for conditions which inflated Type I error rates.